Visit Me, Click Me, Be My Friend: An Analysis of Evidence Networks of User Relationships in BibSonomy

Folke Mitzlaff[1] Dominik Benz[1] Gerd Stumme[1] Andreas Hotho[2]

[1]Knowledge and Data Engineering Group (KDE), University of Kassel
Wilhelmshöher Allee 73, D-34121 Kassel, Germany
{mitzlaff,benz,stumme}@cs.uni-kassel.de

[2]Data Mining and Information Retrieval Group, University of Würzburg
Am Hubland, D-97074 Würzburg, Germany
hotho@informatik.uni-wuerzburg.de

ABSTRACT

The ongoing spread of online social networking and sharing sites has reshaped the way how people interact with each other. Analyzing the relatedness of different users within the resulting large populations of these systems plays an important role for tasks like user recommendation or community detection. Algorithms in these fields typically face the problem that *explicit* user relationships (like friend lists) are often very sparse. Surprisingly, *implicit* evidences (like click logs) of user relations have hardly been considered.

Based on our long-time experience with running the social bookmark and publication sharing platform BibSonomy [4], we identify in this paper different *evidence networks of user relationships* in our system. We broadly classify each network based on whether the links are explicitly established by the users (e. g., friendship or group membership) or accrue implicitly in the running system (e. g., when user u copies an entry of user v). We systematically analyze structural properties of these networks and whether topological closeness (in terms of the length of shortest paths) coincides with semantic similarity between users.

Our results exhibit different characteristics and provide preparatory work for the inclusion of new (and less sparse) information into the process of optimizing community detection or user recommendation algorithms.

Categories and Subject Descriptors: H.4 [Information Systems Applications]: Miscellaneous H.1.m [Information Systems]: Models and Principles

General Terms: Experimentation, Measurement

Keywords: social networks, folksonomies, community detection, user recommendation

1. INTRODUCTION

The participatory nature of many popular Web 2.0 applications like Facebook, YouTube, Flickr, Delicious or Twitter has affected the way how people interact with each other. Catalyzed by the growing availability of mobile web access, the mere fact of being part of an online community bears many instant benefits (like receiving interesting news or staying in touch) for professional and leisure activities.

The resulting "digitalization" of social links into online environments has opened up a number of interesting research questions, many of them originating from the field of social network analysis: *Community detection* is basically concerned with identifying groups of users which share a common interest or expertise, while approaches of *user recommendation* often build personalized models to identify other relevant users within a (possibly large) population. From a more formal perspective, both directions are fundamentally concerned with discovering relationships between users.

Most automated approaches of inferring these links face the problem of judging the quality of their predictions. As a replacement for expensive studies with real users, the algorithms are often evaluated against existing links, as proposed by [15]. This paradigm rewards algorithms which are able to *restore existing structures* (e. g., friendship connections). A typical problem hereby is that these *explicit* links are often sparse or not existent.

Apart from these explict relations, in typical social resource sharing systems one can find a number of *implicit* evidences of user relationships. They comprise, e. g., click-logs or page visit information. In some systems, it is also possible to copy content from other users, which facilitates the assembly of a copy-graph among users. Taking our own system BibSonomy as an example, we identify a set of possible *networks of user relationships* characterized by different degrees of explicitness. Starting with some basic network properties, our first goal is to investigate to which extent these networks exhibit typical characteristics of social networks. We then examine whether topological closeness in these networks coincides with semantic similarity between users. Ultimately, we would like to assess the suitability of each network to serve as a valid surrogate for explicit data.

2. RELATED WORK

Despite the absence of well-established gold-standards, the growing need for automated user community assessment is reflected in a considerable number of proposed paradigms. Evaluation approaches of generated links between users can broadly be divided in content-based and structure-based methods (relying on given links between users).

Ref. [7] introduced metrics for assessing user relatedness and community structure by means of the normalized size of user profile overlaps. They evaluate their metrics in a live setting, focussing on the optimization of the given metrics. Using a metric which is purely based on the structure of graphs, Newman presents algorithms for finding communities and assessing community structure in graphs [13].

Recently Ref. [15] proposed an evaluation technique for recommendation tasks in folksonomies which is based on the reconstruction of existing links (e. g., friendship lists). The performance of a given system is assessed by applying IR-style quality measures. Ref. [14] investigated the relationship of topological closeness (in terms of the length of shortest paths) with semantic similarity between users.

Another aspect of our work is the analysis of implicit link structures which accrue in a running Web 2.0 system. Ref. [2] proposed to present query-logs as an implicit folksonomy where queries can be seen as tags associated to clicked documents. Based on this representation, the authors extracted semantic relations between queries. Ref. [8] analyzed term-co-occurrence-networks in the logfiles of Internet search systems. They showed that the exposed structure is similar to a folksonomy.

Analyzing Web 2.0 data by applying complex network theory goes back to the analysis of (samples from) the web graph [3]. Ref. [10] applied methods from social network analysis as well as complex network theory and analyzed large scale crawls from prominent social networking sites. Some properties common to all considered social networks are worked out and contrasted to properties of the web graph.

3. EVIDENCE NETWORKS

3.1 Preliminaries

We briefly introduce terms and notions used in this paper. For more details, we refer to standard literature [6]. A *graph* $G = (V, E)$ is an ordered pair, consisting of a finite set V which consists of the *vertices* or *nodes*, and a set E of *edges*, which are two element subsets of E. A *directed graph* is defined likewise, with E being defined as a subset of $V \times V$. For simplicity we write in both cases $(u, v) \in E$ for denoting an edge belonging to E and freely use the term *network* as a synonym for a graph.

A *path* $v_0 \to_G v_n$ of *length* n in a graph G is a sequence $v_0, \ldots, v_n$ with $n \geq 1$ and $(v_i, v_{i+1}) \in E$ for $i = 0, \ldots, n-1$. A *shortest path* between nodes u and v is a path $u \to_G v$ of minimal length. The *transitive closure* of a graph $G = (V, E)$ is given by $G^* = (V, E^*)$ with $(u, v) \in E^*$ iff there exists a path $u \to_G v$. A *strongly connected component (scc)* of G is a subset $U \subseteq V$, such that $u \to_{G^*} v$ holds for every $u, v \in U$. A *(weakly) connected component (wcc)* is defined similarly, ignoring the directions of edges $(u, v) \in E$.

For a set V, we denote a *relation* R as a subset $R \subseteq V \times V$. A relation R is naturally mapped to a corresponding graph $G_R := (V, R)$. We say that a relation R among individuals U is *explicit*, if $(u, v) \in R$ only holds, when at least one of u, v deliberately established a connection to the other (e. g., user u added user v as a friend in a online social network). We call R *implicit*, if $(u, v) \in R$ holds as a side effect of u's or v's actions taken.

	Copy	Visit	Click	Follower	Friend	Group
$\|V_i\|$	1427	3381	1151	183	700	550
$\|E_i\|$	4144	8214	1718	171	1012	6693
#scc	1108	2599	963	175	515	90
largest scc	309	717	150	5	17	228
#wcc	37	11	55	37	140	89
largest wcc	1339	3359	1022	83	283	228

Table 1: High level statistics for all relations

3.2 Datasets

For our experiments, we considered three different data sources which accrue in the social bookmark and resource sharing system BibSonomy. Firstly, we used an anonymized dump of all public bookmark and publication posts until January 27, 2010. It consists of 175,521 tags, 5,579 users, 467,291 resources and 2,120,322 tag assignments. The dump also contains BibSonomy's friendship relation among 700 users as well as the *follower* relation, which is explicitly established between user u and v, if u is interested in v's posts and wants to stay informed about new posts.

Furthermore we used BibSonomy's "*click log*", consisting of entries which are generated whenever a logged-in user clicked on a link in BibSonomy, containing the current page URL together with the corresponding link target, date and user name[1]. For our experiments we considered all click log entries until January 25, 2010. Starting in October 9, 2008, this dataset consists of 1,788,867 click events.

We finally considered all available apache web server log files, ranging from October 14, 2007 until January 25, 2010, consisting of around 16 GB compressed log entries. We used all log entries available, ignoring the different time periods, as this is a typical scenario for real life applications.

From all these data sources we extracted *explicit* and *implicit* relations between BibSonomy users[2].

Explicit relations: The *Follower-Graph* $G_1 = (V_1, E_1)$ is a directed graph with $(u, v) \in E_1$ iff user u follows the posts of user v. The *Friend-Graph* $G_2 = (V_2, E_2)$ is a directed graph with $(u, v) \in E_2$ iff user u has added user v as a friend. The friend graph's only purpose is (currently) to restrict access to selected posts so that only users added as "friend" can see them. The *Group-Graph* $G_3 = (V_3, E_3)$ is a undirected graph with $\{u, v\} \in E_3$ iff user u and v share a common group.

Implicit relations: The *Click-Graph* $G_4 = (V_4, E_4)$ is a directed graph with $(u, v) \in E_4$ iff user u has clicked on a link on v's user page. The *Copy-Graph* $G_5 = (V_5, E_5)$ is a directed graph with $(u, v) \in E_5$ iff user u has copied an BibTeX entry from user v. The *Visit-Graph* $G_6 = (V_6, E_6)$ is a directed graph with $(u, v) \in E_6$ iff user u has navigated to v's user page.

Each implicit graph G_i, $i = 4, \ldots, 6$ is given a weighting function $c_i\colon E_i \to \mathbb{N}$ where c_i counts the number of corresponding events (e. g., $c_5(u, v)$ counts the number of posts which user u has copied from v). From all our graphs, we removed spammers and system-introduced artifacts like the artificial user *dblp* or the automatically added tag *imported*. Table 1 summarizes statistical properties of our graphs.

3.3 Structural network properties

[1]For privacy reasons a user may deactivate this feature.
[2]All considered evidence networks are available at `http://www.kde.cs.uni-kassel.de/mitzlaff/papers/2010`

Following [10], we analyze the explicit and implicit user relations given in graphs $G_1, \ldots, G_6$, highlighting properties they share and properties which differentiate them from other online social networks. For reasons of comparability, we restrict our analysis in the following to each of the network's large (weakly) connected component.

Degree distribution: One of the most crucial network properties is the probability distribution ruling the likelihood $p(k)$, that a node v has in- or out-degree k respectively. In most real life networks, the so called *degree distribution* follows a power law [5], that is $p(k) \sim k^{-\alpha}$ where $\alpha > 1$ is the exponent of the distribution. Considering $G_1, \ldots, G_6$, all in- and out-degree distributions except those from the groups graph show a power law like behavior. We omit a further discussion for space reasons.

Link symmetry: Mislove et al. [10] showed for the Flickr, LiveJournal and YouTube that 60-80% of the direct friendship links between users are symmetric. Among others, one reason for this is that refusing a friendship request is considered impolite. Looking at Table 3, we see that BibSonomy's friendship relation differs significantly. Only 43% of the friendship links between users are reciprocal. But as BibSonomy is not used for social networking rather than permission management, the given value is surprisingly high.

When more features are available exclusively along friendship links (e. g., sending posts), the friendship graph's structure will probably change and links will get more and more reciprocal. But concerning the implicit networks we will see, that link asymmetry is determined by a structure common to all our implicit networks.

Path lengths: In many networks it is important to know, whether one can reach a node v from a given node u and the smallest length of all paths connecting v with u if so. Table 3 summarizes for $G_1, \ldots, G_6$ the shortest path statistics. At the first glance, the average shortest path length is strikingly low, ranging from 2.6 to 5.7 hops (considering e. g., to the much larger web graph where in [3] an average of 16.12 hops is given). But many nodes are dead ends, connecting only to a small sub graph. For getting a picture of the connectivity within the graphs, we looked at the fraction of all possible n^2 links in a graph which are present in its transitive closure. The corresponding results are shown in table 3, showing that $G_1, \ldots, G_6$ vary significantly regarding the number of users that are connected by a directed path, though the average shortest paths are approximately the same. In the following we will explore each graph's structure more detailed.

Component structure: Analoguously to the analysis of the webgraph in [3], we analyzed the graph structure relative to a graph's $G = (V, E)$ largest strongly connected component SCC, partitioning V as follows:

$$\begin{aligned} \mathrm{IN} &:= \{u \in V \setminus \mathrm{SCC} \mid \exists v \in \mathrm{SCC}\colon u \rightarrow_G v\} \\ \mathrm{OUT} &:= \{v \in V \setminus \mathrm{SCC} \mid \exists u \in \mathrm{SCC}\colon u \rightarrow_G v\} \\ \mathrm{MISC} &:= V \setminus (\mathrm{IN} \cup \mathrm{OUT} \cup \mathrm{SCC}) \end{aligned}$$

In [3] the webgraph was shown to be partitioned in approximately evenly sized node sets aligned around a strongly connected component which covered 27% of the webgraph's node set. For Flickr and YouTube the largest strongly connected componets, covering more than 40% of the corresponding node sets were identified [10].

We saw in the last section, that $G_1, \ldots, G_6$ varied in the

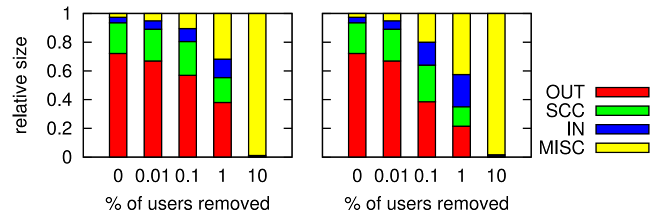

Figure 1: Gradually removing central users from the visit graph, ordered descending by in-degree (left) and ordered descending by out-degree (right)

	Copy	Visit	Click	Follower	Friend	Group
|SCC|	309	717	150	5	16	228
|IN|	264	128	146	13	94	0
|OUT|	525	2423	514	14	17	0
|MISC|	241	91	212	51	156	0

Table 2: Core structure

fraction of all possible links between nodes which are present in the corresponding transitive closure. Table 2 explains the differences. Graphs which are clearly structured in IN, SCC and OUT allow more node-to-node paths as those, where a significant larger part of the nodes is located in the MISC set.

But Table 2 reveals also, that in all implicit networks the OUT set is significantly larger than the IN set. The proportion $\frac{|\mathrm{IN}|}{|\mathrm{OUT}|}$ varies largly, ranging from 0.05 (visit) to 0.5 (copy). Especially the visit graph is clearly structured by the SCC, where only 3% of the nodes do not lie on a walk from a node $u \in \mathrm{IN}$ to $v \in \mathrm{OUT}$. Most probably the SCC in the implicit graphs is densely structured around BibSonomy's most active users. For analyzing the structure of a strong component, we conducted an experiment similar to [3] and [10], where central nodes from the webgraph and some social networks were gradually removed (sorted in descending order by node degree). Figure 1 shows representatively for all implicit graphs, that the structure is rapidly destroyed and completely vanishes after removing the 10% of the most active users (quantified by out- and in-degree respectively). Caution is necessary when interpreting Figure 1, as it only reflects the proportions of the graph. Removing the top four users results in removing half of the links and half of the users. This is due the fact, that these top users were the system maintainers, responsible (among others) for spotting spammers which results in a very high out-degree. Notably the overall structure was sustained, which indicates that it is due to underlying usage patterns. The explicit graphs are not structured accordingly. Notably, the friend graph again shows its asymmetric nature, having a significantly larger IN than OUT set, which corresponds to users having added asymmetric friend relations.

Link degree correlations: A phenomenon often observed in social networks is that nodes tend to connect with other nodes of "the same type". This selective linking pattern is called "assortative mixing" or "homophily" [12]. In most social networks, assortative mixing relative to the node degree is observed [11].

The *degree correlation function* k_{nn} maps a node's out-degree value k to the mean in-degree of all nodes incident to a node with out-degree k [16]. A simple plot of k_{nn} with logarithmic scale reveals fundamental tendencies: If nodes with

	Copy	Visit	Click	Follower	Friend	Group
APL	4.31	3.87	4.79	2.61	3.5	2.98
Radius	1	1	1	1	1	4
Diameter	15	11	14	7	10	7
FL*	27%	24%	19%	7%	7%	100%
SL	8%	12%	12%	11%	43%	100%
SL*	20%	19%	12%	6%	15%	100%

Table 3: Link properties, including average path length (APL), fraction of links in transitive closure (FL*), fraction of links which are symmetric (SL) and fraction which are symmetric in the transitive closure (SL*)

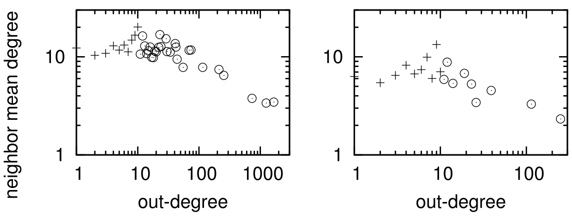

Figure 2: Degree correlation function k_{nn} for the visit (left) and click graph (right).

larger degree connect to other high degree nodes (assortativity), the plot follows asymptotically an ascending line. If the plot shows an descending tendency, nodes with high degrees tend to connect with nodes of lower degree (dissortativity).

Figure 2 shows representatively a plot of k_{nn} for G_4 and G_6. For all implicit user relation graphs the degree correlation functions share the same pattern, suggesting assortativity up to a small degree (cross) but then dissortativity (circle). Due to sparsity of the data (there are few different node degrees), it is not possible to say whether this observation is an artifact (e. g., due to the limited size of the graphs) or systematic. But it is worth mentioning, that it was also observed within the network of scientific collaboration [1] as well as in [4] and [8], where respectively BibSonomy's underlying folksonomy and the network of term co-occurence in search queries were analyzed. In these works, both regions were treated separately, concluding that assortativity was present, as the other part of the graph was rejected due to its limited size.

3.4 Semantic Structure

The analysis of the last section has focussed on several inherent network properties of each analyzed evidence network of user relationship. A key insight was that from a graph-theoretic perspective, the networks exhibit different characteristics, and none of them can be understood as a "typical" social network. In this section we will go one step further and take into account information which is not present in the networks itself — namely background information about the *semantic profile* of each node. Despite the differences to a typical social network reported above, it is a natural hypothesis to assume that, e. g., two users which are close in the click network can be expected to share some common interest, which is reflected in a higher "semantic similarity" between these user nodes. In this way we establish a connection between structural properties of our networks and a *semantic* dimension of user relatedness.

Here we also face of course the problem of measuring the "true" semantic similarity between two users. Here we build

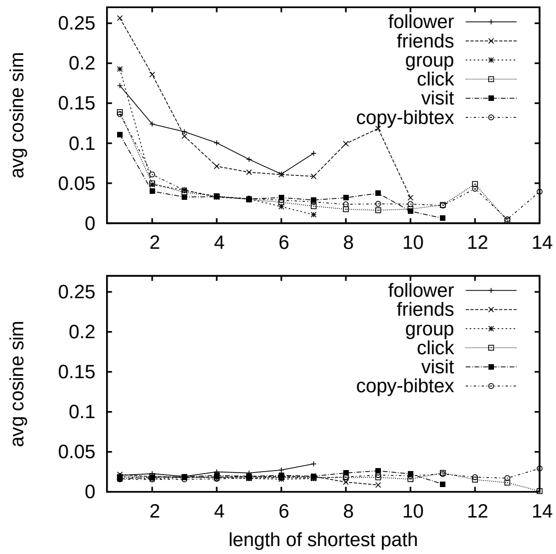

Figure 3: Average Cosine Similarity (y-axis) between tagclouds of users as a function of their distance in each of our networks. The upper figure represents the computation on the original folksonomy, the lower figure the null model created by reshuffling the tags for each user.

on our prior work on semantic analysis of folksonomies [9], where we discovered that the similarity between tagclouds is a valid proxy for semantic relatedness. We compute this similarity in the vector space $\mathbb{R}^T$, where, for user u, the entries of the vector $(u_1, \ldots, u_T) \in \mathbb{R}^T$ are defined by $u_t := w(u,t)$ for $t \in T$ where $w(u,t)$ is the number of times user u has used tag t to tag one of her resources. Each vector can be interpreted as a "semantic profile" of the underlying user, represented by the distribution of her tag usage. Similarity is computed by the cosine similarity (refer to [9] for details).

Inspired by the presentation in [14], we plot the average semantic similarity between all pairs (u, v) of users (as obtained by the cosine similarity in the tag vector space) against the shortest path between u and v in each of our networks. This is done in Figure 3 (upper figure).

The first obvious observation is that the highest average *semantic* similarity is found for the smallest *topological* distance of 1. With growing lengths of the shortest paths (x-axis), the average semantic similarity decreases quickly towards 0. This holds for almost all measures. Manual inspection of the peaks at greater distances in the friends and follower networks showed that they are probably due to the sparsity of our data.

In summary, these results indicate a correlation of topological proximity and semantic similarity between the nodes in each of our observed networks. This confirms in a more formal way that shared interests between users are reflected in a higher degree of interaction between them — in an implicit or explicit manner.

Following again the ideas of [14], we also want to ensure that the obverved phenomena are not due to statistical effects of, e. g., users with very large tagclouds in our networks being more similar to others. To this end, we created the null model proposed in [14] and reshuffled the tags for each user, while keeping the original (global and per-user) frequencies. For this shuffled folksonomy, we recomputed the pairwise semantic similarity and plotted it again against the lengths of the shortest paths in each of our networks. The result can be seen in Figure 3 (lower figure). One can see that the shuffling process has eliminated completely the correlation observed before (upper part of Figure 3); this means that the latter is a real and not a statistical effect.

Interestingly, this effect seems to be stronger for the explicit networks: The highest average *semantic* similarity is found for the friends and the follower network. The click and copy network are in between, and the "most implicit" visit network yields the lowest values, while still exhibiting the same peak for lower distances as the other networks. This ranking is in line with an intuitive "quality" judgement of the underlying activity: The more explicit the networks are, the better they mirror semantic closeness of users. However, it is important to notice that the less explicit networks are of course much bigger and more dense; hence they can still be of great use in complementing the sparse explicit structures.

4. CONCLUSIONS

In this work we analyzed and characterized a number of *evidence networks for user relationships* which have evolved within our system. Our primary motivation hereby was to inform a larger research community about the availability of this kind of data, and about the individual characteristics of the networks induced by users visiting, clicking, copying from each other, and being each others friends, followers or group members. After a first broad distinction of these graphs into more *implicit* and more *explicit* structures, we analyzed several network properties and examined the correlation of topological proximity and semantic similarity between nodes in the respective networks. Our results confirm the intuitive assumption that more explicit networks mirror more closely semantic relationships between users; however, given the larger size and higher density of implicit networks like the visit or click graph, these can still be considered as a valuable indicator of shared interests.

We do *not* want to argue that considering explicit structures alone (as typically done in the literature, especially in the field of user recommendations) is not enough. However, our results indicate that one has to pay attention to the way how the social networking facilities are embedded in the system. While one would expect that the "friend"-relationship within BibSonomy resembles probably most closely social links, our analysis showed that the resulting graph only partially exhibits the characteristics of a social network. This might be due to the fact that the friendship-feature of BibSonomy is focussed on permissions (one can set the visibility of certain entries to "friends only") rather than on, e. g., networking. By this we would like to exemplify that the context of the emergence of a specific evidence network has to be taken into account when including it into a task like community detection or user recommendation.

We are aware that this is work has a preparatory character towards a more detailed analysis of the interplay between each of our networks, especially in the context of a given task. We see our results as a stimulation for social network researchers to recognize the presence and value of *implicit* evidences of user relationships. Futhermore, it is our goal to share the unique possibility of posessing this broad variety of user interaction logs with a broader research community.

Acknowledgement. This work has been partially funded by the VENUS and COMMUNE projects, funded by Land Hessen and Hertie Foundation, respectively.

5. REFERENCES

[1] J. A. Almendral, J. Oliveira, L. López, J. Mendes, and M. A. Sanjuán. The network of scientific collaborations within the european framework programme. *Physica A: Statistical Mechanics and its Applications*, 384(2):675 – 683, 2007.

[2] R. Baeza-Yates and A. Tiberi. Extracting semantic relations from query logs. In *Proc. of the 13th ACM SIGKDD conf.*, page 85. ACM, 2007.

[3] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33(1-6):309–320, 2000.

[4] C. Cattuto, C. Schmitz, A. Baldassarri, V. D. P. Servedio, V. Loreto, A. Hotho, M. Grahl, and G. Stumme. Network properties of folksonomies. *AI Communications Journal*, 20(4):245–262, 2007.

[5] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data, 2007.

[6] R. Diestel. *Graph theory.* Springer, Berlin, 2006.

[7] P. N. Karamolegkos, C. Z. Patrikakis, N. D. Doulamis, P. T. Vlacheas, and I. G. Nikolakopoulos. An evaluation study of clustering algorithms in the scope of user communities assessment. *Computers & Mathematics with Applications*, 58(8):1498–1519, 2009.

[8] B. Krause, R. Jäschke, A. Hotho, and G. Stumme. Logsonomy - social information retrieval with logdata. In *HT08: Proc. of the 19th ACM Conf. on Hypertext and Hypermedia*, pages 157–166. ACM, 2008.

[9] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stumme. Evaluating similarity measures for emergent semantics of social tagging. In *18th Int'l WWW Conference*, pages 641–641, 2009.

[10] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *7th ACM SIGCOMM conf. on Internet measurement*, page 42. ACM, 2007.

[11] M. Newman and J. Park. Why social networks are different from other types of networks. *Physical Review E*, 68(3):36122, 2003.

[12] M. E. J. Newman. Mixing patterns in networks. *Phys. Rev. E*, 67(2):026126, Feb 2003.

[13] M. E. J. Newman. Detecting community structure in networks. *The European Physical Journal B*, 38, 2004.

[14] R. Schifanella, A. Barrat, C. Cattuto, B. Markines, and F. Menczer. Folks in folksonomies: social link prediction from shared metadata. In *Proc. 3rd ACM Int'l Conf. on Web search and data mining*, pages 271–280, New York, NY, USA, 2010. ACM.

[15] S. Siersdorfer and S. Sizov. Social recommender systems for web 2.0 folksonomies. In *HT09: Proc. of the 20th ACM Conf. on Hypertext and Hypermedia*, pages 261–270, New York, NY, USA, 2009. ACM.

[16] A. Vázquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the Internet. *Physical Review E*, 65(6):66130, 2002.